DOCUMENT RESUME

ED 280 853 TM 870 164

AUTHOR Chelimsky, Eleanor
TITLE What Have We Learned about the Politics of Program Evaluation?
PUB DATE Oct 86
NOTE 40p.; Paper presented at the Annual Meeting of the American Evaluation Association (Kansas City, KS, October 21-November 1, 1986).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Decision Making; Educational Research; Evaluation Criteria; *Evaluation Methods; Evaluation Problems; *Evaluation Utilization; Evaluators; Federal Government; Models; *Policy Formation; *Politics; Politics of Education; *Program Evaluation; Research Design

ABSTRACT

The politics of program evaluation are discussed from the personal perspective of the Director of the General Accounting Office's Program Evaluation and Methodology Division, which has produced reports for committees of the United States Congress. It is concluded that successful evaluations must be useful to others and must understand the political system in which evaluation operates. The political system encompasses not only executive policymakers, but also the legislative (and, perhaps, in the future, the judicial) branch, and their interactions. Evaluation planners should understand that the policy question must be of interest to the user and that the findings must answer that question. Evaluations must be credible, timely in accordance with the policymaking cycle, and cost effective. A good relationship with the executive agency implementing the evaluation's recommendations is significant. It is also suggested that evaluative thinking be introduced into the policy forum before legislation is proposed. Panels of experts strengthen an evaluation's credibility. It is important that the evaluator master a wide variety of approaches and that the user understand the evaluation design process. Editing of the evalution report, prioritization of findings, and meta-analysis or evaluation synthesis are useful. Impact questions are generally asked by legislators, but descriptive questions are not typically raised by policymakers. Taking political processes into account affects time allocation; more time must be devoted to negotiation, discussion, briefing, accuracy-checking, prioritization, and presentation. Traditional evaluation has improved as a result of the major modifications described in this presentation. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED280853

WHAT HAVE WE LEARNED ABOUT THE POLITICS OF PROGRAM EVALUATION?

Plenary Address to the
American Evaluation Association
October 30, 1986
in
Kansas City, Missouri

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E Chelimsky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Eleanor Chelimsky
Director, GAO's
Program Evaluation &
Methodology Division

Note: The views and opinions expressed by the author are her own and should not be construed to be the policy or position of the General Accounting Office.

TM 870 164

Good morning, everyone. It's a delight to be here, to see you all flourishing, and to get a sense of this society of evaluators continuing to grow in size and in accomplishments despite the inhospitable terrain in government today. I see us as being a little bit like suburban dandelions: you know . . . people spend their whole weekend trying to get rid of them but when Monday afternoon rolls around, there they are back again in full and glorious bloom: bright, stronger than ever, and not just unabashed, but thriving.

This meeting, I know, marks the point at which, as a society, we've begun to think about our past, about the paths we've travelled and our reasons for choosing them, and about the progress we've made, all along the line. I've been asked, in particular, to say what I think we've learned about the politics of evaluation.

To do that, as you might well imagine, I went back in time -- oh, say 10 years -- to remember what we used to say or write about the issue of politics in evaluation and then contrasted that with our approaches today. Here, of course, you should note that I'm speaking from my own personal perspective and experience as head of the Program Evaluation and Methodology Division (PEMD) at GAO, and I'm also speaking about a process -- the political or policy process -- that is so enormously variegated, so full of chiar-oscuro, so inclusive and tolerant of exceptions and special cases, that generalizations hold some peril for the generalizer and should be understood within the context of this experience and this environment. The experience, however, is not insignificant. PEMD has now produced more than 50 reports for different committees of

the Congress and we have also worked, in a more transient capacity, with many more.

What then do we mean when we talk about the "politics of program evaluation"? David Easton wrote 30 years ago that "the study of politics is concerned with understanding how authoritative decisions are made and executed for a society" (Easton, 1957, p. 383), and that "the output of a political system is a political decision or policy." (Easton, p. 395) Public policy, then, is the product of politics. Evaluation, with its purpose of providing high-quality information to decisionmakers, thus automatically claims a role for itself in the political process, based on the idea that the best information, made available to decisionmakers, would surely be useful to them in making and executing public policy. Unfortunately, in the very early 70's, rumors began to be heard in the land (later, alas, confirmed by studies) to the effect that evaluations were not, in fact, often used in decisionmaking.

At first we tried to come to grips with the problem, in our best Kuhnian fashion, by looking around for ways out of the difficulty that wouldn't require too many major changes in our paradigm or too many shifts in our workstyle. We said that if our work wasn't used, this was because some evaluations simply weren't good enough (Bernstein and Freeman, 1975), forgetting -- or not yet realizing -- that policymakers typically use any data at hand (Stromsdorfer, 1981, pp. 1 and 8) and that some very mediocre evaluation work has in fact been quite well used. Or we said that maybe evaluation <u>was</u> used, but you simply couldn't trace that use because evaluation was just one of many factors influencing

decisionmaking (Chelimsky, 1985, p. 16). We said that maybe "use in decisionmaking" was too narrow a goal for evaluation, that there were many meanings -- at least seven -- attributable to the term "research utilization" (Weiss, 1979, in Chelimsky, 1985, pp. 203-212), and we said that, in any case, what we needed to develop was information supporting negotiation rather than information supporting decisionmaking (Cronbach, 1980, p. 4, thesis 26). What we were doing, of course, was to turn the question of "why isn't evaluation used?" into the question, "what, in fact, is reasonable use?" This in turn led to the idea that expectations for the use of evaluations were perhaps unrealistically high (Cronbach and Associates, 1980, p. 5).

However, there had also been a 1973 paper by Carol Weiss, entitled Where Politics and Evaluation Meet, in which she made what, for me, were two crucial points: first that "as a matter of record, relatively few evaluation studies have had a noticeable effect on the making and remaking of public policy" (Weiss, 1973, p. 40), and second, that "only with sensitivity to the politics of evaluation research can the evaluator be as strategically useful as he should be" (Weiss, 1973, p. 38). Thus, the lack of evaluation use was beginning both to be recognized as such and to be seen not so much as a problem that could be whisked away without major changes in the evaluation paradigm, but rather as the newly perceived and very difficult problem of integrating the disparate worlds of politics and of evaluation research.

After the appearance of this paper, and between about 1975 and 1980, a great spate of articles and studies were published that

sought to relate evaluations generally to the larger political and policy context within which they operate. One of the most important of these was certainly the 1980 book by Lee Cronbach and Associates, entitled Toward Reform of Program Evaluation. It is in looking back today at many of these studies, I think, that we can get the best sense not only of what we have learned about politics and evaluation, but also, how we have changed in our outlook and practice.

While it was quite typical in those studies to read statements attesting to the importance of politics to evaluation, statements like: "A theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts" (Cronbach, p. 3, thesis 11), or "The evaluators' professional conclusions cannot substitute for the political process," (Cronbach, p. 3, thesis 13), there were not many cogent, action-oriented, systematic and specific discussions of how the integration of evaluation and politics should or could take place. We did hear suggestions for the introduction of "research brokerage" (Sundquist, 1978), or of "a bridging function," (Chelimsky, 1977, Vol. I, p. 117), both of which sought to solve the integration problem by interposing some person or unit between the evaluation and the political user. But this suggestion required a whole new breed of "interface analysts" and it never bore fruit, as James Sundquist predicted when he broached the idea in 1978.

Perhaps one reason why concrete suggestions were so hard to come by was that the political domain with which evaluation had to interact was so little understood by evaluators. I remember that

in 1975, when I was organizing the MITRE Corporation's symposium on the use of evaluation, it was considered quite unusual to have evaluators from many different evaluative disciplines (such as, psychology, sociology, economics, political science, or operations research) and from many different substantive program areas (not only the usual health, education, and welfare, but also, for example, energy and the environment), it was, as I say, quite a novelty for these evaluators not only to sit down <u>together</u>, but to sit down with their policymaking and program-managing sponsors and users in different federal agencies. What came out of that symposium was a set of concrete user dissatisfactions with evaluation's performance, many of which were political, and many of which were quite shocking to the evaluators present.

We heard that evaluations often can't be used because of bureaucratic relationships and conflicts (Chelimsky, 1977, Vol. I, pp. 35-36), that some program managers really <u>prefer</u> to be ignorant (Chelimsky, 1977, Vol. I, p. 46), and that evaluations will only be used if management really wants them (Chelimsky, 1977, Vol. I, pp. 145 and 151). We heard that evaluation seemed to its users to be an ivory-tower process (Chelimsky, 1977, Vol. I, p. 36), and that when it emerged from that tower, it was too late to be useful, too full of jargon to be understood, too lengthy for the user reading-time available, and too prone to be answering a question quite different from the policy question originally posed (Chelimsky, 1977, Vol. II, pp. 33 -44). All of this was important to hear, not only because it was a kind of beginning excursion into the agency politics of evaluation, but also because the information had a

chance of being assimilated or internalized by evaluators, given that most evaluations at that time had been developed for agencies of the executive branch.

Indeed, when Cronbach's book on evaluation reform came out in 1980, it was clear that many of the lessons of agency politics learned at the MITRE symposium had in fact been integrated by at least some evaluators. But we were still far from having a political framework comprehensive enough to plan our integration work against, because, for the most part, legislative politics had remained unexamined.

This is not a criticism of Cronbach's book or of the MITRE symposium, but rather a reminder that time passes, experience broadens, and we learn new things. I think it was Galbraith who said that each step in a learning process builds on the one before and that critics of the earlier steps are not wiser, only later. But, while Cronbach properly reproached us in 1980 with taking a much too narrow view of the political arena in which agency decisions are made and evaluation use occurs (Cronbach, pp. 54-62), this view, in turn, seems narrow to us today because it failed to include legislative politics.

Now this is not just a matter of leaving something out. One problem is that many points made, which are right on target when it comes to agency politics, are notably inappropriate with regard to legislative politics. For example, after considering what we've now learned in 6 years of working with the Congress, it seems clear we need to rethink premises like "a social agenda almost never calls for a choice between fixed alternatives" (Cronbach, p. 62);

or "only rarely are 'go/no go' decisions made"(Cronbach, p. 62); or "timeliness is a much-overrated concern" (Cronbach, p. 63). The Congress, after all does indeed continually make choices between fixed alternatives and does vote on go/no go decisions; and an evaluator who didn't understand the critical importance of the timing of congressional policy cycles -- involving program authorization, reauthorization, appropriations, and budget -- such an evaluator, as I say, would be spinning his or her wheels and wasting a lot of money. Indeed, I would submit that the importance of timeliness with regard to legislative politics, at least, cannot be over-stated (Florio, Behrman and Goltz, 1979, pp. 68-69). A second problem is that, without an understanding of legislative politics, we cannot understand -- or worse, we actively misunderstand -- the dynamics of cross-branch politics: that is, the reciprocal process of legislative/executive agency interactions.

The whole point, in fact, shows how our understanding of politics has progressed. Whereas Cronbach properly warned us of the dangers inherent in over-generalizing what is true for one type of evaluation to all evaluations (Cronbach, p. 51), here I would, in my turn, caution against over-generalizing from agency evaluation politics to all evaluation politics.

Another result of failing to include legislative politics in our thinking, I believe, is the way we continue to restrict evaluation's application to social issues only. Over and over, in the literature, we read of evaluation's preoccupation with social programs, with social reform, with social concerns. But, of

course, legislative politics means that evaluation must broaden its focus to be optimally useful. Working for the Congress, evaluators are expected to exercise their skills in all policy areas, social or not. Indeed, it was because I had included a single legislative perspective on evaluation in the MITRE symposium that I recognized the need to also include energy and the environment -- two largely non-social, technological areas that were and are of great concern to the Congress -- in the 1976 symposium agenda.

If I had to characterize the politics of evaluation today, taking into account both executive and legislative branch politics as we've learned to know them and their interactions with each other as well as with evaluation, I would say that what we are dealing with overall is a process of continuous translation for which the evaluator must assume a major share of responsibility. I would represent the process as a five-part continuum which involves, first, the development of a <u>policy</u> question. (Let me just note here in passing that when I talk about policy questions, I do not exclude program questions, or imply that all questions for evaluation are policy questions, but rather, I include all types of questions under this rubric insofar as they are asked by policy actors and emerge from the political process). This policy question is then translated into an <u>evaluation</u> question which typically alters the policy question so that it becomes researchable. Then, in the third part of the continuum, the evaluation question is translated into the evaluation proper: its design, its performance and the reporting of its findings. The fourth part is the translation back into policy language of the results of the

evaluation; that is; it is the policy answer to the original policy question. Finally, the fifth part comports the use of the evaluation findings somewhere in the cross-branch political process, and the return, completing the continuum, to the formulation of new policy questions.

Now what does such a framework imply in terms of changes from the old evaluation paradigm? Let's examine the five parts of the continuum and see what differences they, and the framework as a whole, may make for the way in which evaluators do evaluation. I would like to discuss 14 points of difference that I think are important.

Part I: The Development of the Policy Question

A first point, then, is that, if we use this framework, the evaluation can no longer be seen as completely a creature of the evaluator's choosing -- with the evaluator being entirely free to select the program he or she wants to evaluate (Weiss, 1973, p. 42), and to determine alone the questions to be asked about the program (Cronbach, p. 5, theses 30 and 37). Instead, insofar as we are dealing with evaluations that have been sponsored and are intended to be used in the political environment, the framework means that the choice of the program to evaluate emerges in real terms from the political process, with the determination of the types of policy questions to be asked a function of the decisionmakers, whether legislative or executive or both. (It does happen that both congressional and agency policymakers are in accord about what the important questions are, as in this year's

emphasis on teenage pregnancy, for example, or last year's issue of cost containment in health care services.)

As such, this framework involves both an empowering of the decisionmaker with regard to the evaluation product, and a shift in the role of the evaluators from the political one of seeking to reform society, or "to improve the way that society copes with social problems" (Weiss, 1973, p. 42), to the scientific one of bringing the best possible information to bear on a wide variety of policy questions. Such a change squares well with definitions of evaluation that see it as "the application of scientific research methods to the assessment of program concepts, implementation, and effectiveness" (USGAO, 1984, p. i) or as the provision of "careful and unbiased data" (Weiss, 1973, p. 37) about programs, or as "a process by which society learns about itself" (Cronbach, p. 2, thesis 1). It does not square well with the notion that, "an evaluator should not undertake to serve an agency unless he is in sympathy with its general mission" (Cronbach, p. 211), or with program advocacy or partisanship definitions of evaluation in which evaluators seek to be a voice for the program, or a voice for the underdog, or for some other special group. Similarly, it cannot accommodate precepts like "it is unwise for evaluation to focus on whether a project has 'attained its goals'" (Cronbach, p. 5, thesis 30), first, because such a precept assumes that it is the evaluator who chooses the questions, and second, because the politics of the legislative oversight function may require that evaluators focus on precisely that question.

A second point is that the primary purpose of having the policy question be one that is really of major concern to politicians is to maximize the use of the findings. When the user generates the questions and when the purpose of the evaluation is to produce the highest quality information for policy or program use, this has the effect of eliminating any congressional suspicions that "the questions set for discovery (by the evaluators) have predetermined answers" or that "the assumed posture of objectivity among program evaluators masks subtle but important biases and hidden agendas" (Chelimsky, 1977, Vol. I, p. 68). What I am suggesting is that, in legislative politics, the most credible evaluations are those whose use will be maximized. Now this may seem somewhat paradoxical, given that I've just said that high methodological quality may not ensure use and when we have seen that adversaries in the political process will often use any data that support their case (Stromsferfer, p. 8). Indeed, some people believe (Lindblom, 1968) that the best way to achieve policy use is through partisan evaluation that supports either the policymaker's ideology or the evaluator's own. But credibility involves more than methodological quality: it involves responsiveness to the specific policy question and information need. And partisan use by policymakers is not helpful for the credibility of an evaluation. On the contrary, partisanship is, at best, a short-term strategy for use which is not well adapted to legislative politics. Indeed, in the Congress, partisan work that is well received by one committee would be quickly discredited by another; and worse, the reputation for partisanship would remain with the evaluator over

the longer term. Independent analysis that is strong methodologically and seeks to be as objective as is humanly possible, serves a much wider legislative audience and can be the focus of cross-committee or even House- and Senate-wide debate. (We have actually seen this happen on three separate occasions now in our work on chemical warfare.) I'm not trying to make the case that an evaluation which is read and whose findings are used in the policy process will not have a political influence. Of course it will: such an influence is inherent in the fundamental interaction of evaluation with politics. But evaluation's main value to policy in the long run is not its capacity for political influence but its contribution of systematic, scholarly, independent, critical thinking to the decisionmaking process.

A third point regarding the political origin of the policy question is the concern that decisionmakers may fail to ask certain kinds of questions either because they don't want to, or because they don't think of them, or because there isn't time to get answers. In a paper I wrote recently (Chelimsky, in press), I explored some of the questions decisionmakers at different levels and positions in both branches of government do tend to ask, and pointed out that there are some areas in which evaluators are not often asked questions despite the fact that they could probably contribute powerfully to useful answers. An example of this is the rarety with which decisionmakers pose relevant evaluative questions prior to the introduction of a new program. As a result, there is habitually a serious dearth of evaluative information entering the area of policy formulation in time to influence the proposal.

Many of us spend much of our time doing ex-post, retrospective studies, and it is true that these are and will continue to be the meat and potatoes of evaluation. The Congress asks us for them and asks the executive branch to do them, they are needed and so of course they are a major part of the evaluation menu. But these studies are not the easiest ones to insert into the political process and they may well be the least propitious from the viewpoint of use. When we at GAO do one of them, we always try to focus hard on what it will be possible to accomplish uniquely through legislative means, because we know how hard it will be to change much of anything in a long-established, well-entrenched agency program.

By contrast, before a program has started, evaluators can have an enormous effect in improving the reasoning behind program purposes or goals, in identifying the problems to be addressed, in selecting the best point of intervention and the type of intervention most likely to succeed. The difficulty here is the tempo at which new programs are sometimes introduced. Just recently, Senator Moynihan complained that the provisions of the Gramm-Rudman-Hollings bill were still being tinkered with while he was being asked to vote on it on the floor. The pace often becomes so frantic that the lead-time necessary to gear up for evaluative work that can be ready soon enough to be useful, is simply impossible to obtain.

At GAO we're now developing a method I call the Evaluation Planning Review which is specifically intended to be useful in new program formulation. We have just given it a first try by looking

at a proposed program focusing on Teenage Pregnancy (USGAO, 1986). Essentially, the method seeks to gather up information on what is known about past, similar programs and apply the experience to the architecture of the new one. Senator Chafee asked us to look at the bill he was introducing, we managed to secure four good months to do the work, and I think it's been a major success both from the legislative point of view and our own. From a more general, political perspective, it seems to me that providing understanding ahead of time of how a program might work can render an invaluable public service either by helping to shore up a poorly thought-out program, or by validating the basic soundness of what is to be undertaken. So yes, there are questions that decisionmakers don't pose to evaluators that could usefully be posed, which seems, a priori, to be a problem for the framework. However, first, even when evaluators have been free to choose the questions, this particular type of question has not often been asked. Second, evaluators can always influence the next round of policy questions through their products. I'll return to this issue in a little while.

Part 2: The Translation of the Policy Question Into an Evaluation Question

Now let's turn to the second part of the continuum, which is the translation of the policy question into an evaluation question. We have found in working with the Congress that this is one of the most sensitive and important political interactions in the entire process. So a fourth point about how the framework implies changes from the old evaluation paradigm is that the move from policy

question to evaluation question has to be given a great deal of thought, care, and attention. This is where the sponsor is going to hear whether the policy question being asked is researchable or not. If the evaluator concludes that it is not researchable, this may be because the issue or program is immature; for example, it may be impossible to answer cause-and-effect questions if there has been little or no prior conceptual development in the field to guide the execution of such work. In this case, it would be wise to try to transform the question from the cause-and-effect type to another type that could be answered, if in fact such an answer would also be useful to the decisionmaker. One could move, say, from asking what the factors are that cause high quality and productivity in an organization, to asking either what is known about high quality in an organization and how it is defined, or what the characteristics are of people and work processes in a number of selected organizations whose high quality and productivity are recognized. The point I want to make here is that this eminently rational process of transforming a cause-and-effect question into a descriptive question <u>based on evidence of researchability</u> is fraught with risks.

To begin with, sponsors don't like to have their questions modified; to them, this smacks of the researcher substituting his or her wisdom for that of the policymaker. So a change needs to be very carefully and persuasively explained and negotiated. Then again, in the Congress, there is often a real need for findings that are generalizable to the nation, whether or not a national study is feasible in terms of time, costs and the type or locus of

the program. In programs heavily dependent on state variation, for example, there is a continuing problem of how to obtain in-depth knowledge that can help illuminate that variation, and at the same time develop more superficial but politically needed knowledge about what the situation is nation-wide, without also bankrupting yourself. So the translation is not always an obvious or easy one. Finally, in thinking forward to Part 4 of the framework continuum -- when evaluation findings will have to be translated back into policy findings -- it's critical for both the evaluator and the sponsor to remember the importance that a powerful methodology will have at that time for being persuasive, if the evaluation deals with a controversial policy area. That is, it is rarely prudent to enter a burning political debate armed only with a case study, even if that case study were the most rational approach to the original policy question.

So the movement from policy question to evaluation question must be prepared very carefully, with a lot of thought given to the sponsor's time requirements, the cost, the relation of the type of findings expected to the particular policy need, the kind of program and its locus, the prior research done, and the likely controversiality of the issues.

A fifth point is that the evaluator also needs to think beyond the sponsor, in this part of the continuum, to the entire federal context of the program. That is, the question may have been raised by legislative policy actors, but in order for any evaluation recommendations to have an effect, the evaluator knows it will be necessary for executive branch policy actors to think those

recommendations are right and to act on them. This then entails another sub-complex of questions for the evaluator who wants the evaluation to be useful, including such issues as, Have we got the right problem from the agency viewpoint? If they don't think so, is that because it's unimportant or because they just don't want to think about it? Is the question perhaps symptomatic of a deeper problem that should be addressed? Who is in a position to do something about the problem? What if a power change occurred during the conduct of the evaluation -- let's say a new administration came in or the control of the Senate changed hands -- what effect would that have on the usefulness of the findings? It is not that the evaluator can necessarily do something about all of these issues. But they and others may have a great impact on the eventual use of the findings. At minimum, they mean that legislative evaluators have a major interest in maintaining good relations with executive branch agency evaluators if it is those agencies which must implement the recommendations and ultimately resolve the problem being treated by the evaluation.

A related consideration in conducting the negotiations on the change from policy to evaluation question, involves remembering the impatience of policymakers and their staffs with the typically equivocal nature of findings. For them, what is uppermost is their political need to take unequivocal policy positions. This re-emphasizes the importance for evaluators working within a policy context to be very aware of the type of political debate into which their finished evaluation will likely enter, and to structure the evaluation design -- once the evaluation question is settled -- not

so as to deal with every conceivable issue, but to produce strong, well focused information which will be usable in that particular debate.

Part 3: The Translation of the Evaluation Question Into the Evaluation Proper

The third part of the continuum involves doing the evaluation. Although the evaluator now begins a phase of the work in which he or she is in at least relative control, both the evaluation design and the writing of the report need to be understood with the political context in mind. So a sixth point about the changes in evaluation practice that are brought about by the use of this framework has to do with the narrow limits of policymaker interest in past experience. Decisionmakers look forward; evaluators look back. Although it is often the case that policy actors are genuinely interested in knowing what effects a program had, their interest almost never stops there. Almost invariably the evaluator will be asked to say what the evaluation findings mean in prescriptive terms: if it is a descriptive study, there'll usually be a normative question posed and, nearly always, an invitation to say what the findings signify for future programs. In PEMD we have finally grasped that we must expect these kinds of questions and even in an exploratory study, if it's for the Congress, we try to build in a panel of experts, very carefully chosen to reflect the widest possible spectrum of political opinion as well as both substantive and methodological expertise. We bring these experts in early in the study to help us develop the criteria we know we'll need to make use of in answering the normative questions that will

be asked. For example, in a descriptive management study, how will we decide what degree of communication in an organization is adequate, what degree is inadequate? In a medical study, what counts as "progress," what counts as "lack of progress" in treatment or patient-management techniques? In a study of research utilization, how many omissions or distortions of the findings does it take to constitute misuse? Cronbach wrote in 1980: "An evaluation of a particular project has its greatest implications for projects that will be put in place in the future" (Cronbach, 1980, p. 7, thesis 51). What we have now learned is how to build features into a design that will allow us to speak a little more strongly about the significance of our findings for those future projects.

A seventh point of change implied by the framework for the execution of the evaluation design involves the need for a very wide range of approaches and methods, often used in combination so that the strengths of one can palliate the weaknesses of another. Responsiveness to a policy question doesn't allow the luxury of "chapels dedicated to the glorification of particular styles" (Cronbach, p. 1). However, the political environment has its exigencies even here. National data are wanted, nearly always. Again, there is a similar kind of mystique surrounding heavily quantitative impact evaluations; for congressional staff generally, they are far and away the most highly considered studies. (A 1979 survey showed that "information on the impact of legislative policy and authorized programs is significantly greater in importance" than other types of information, such as demographic information on

populations served and affected, opinions and reactions of interested population groups, or program descriptive information, Florio, Behrmann and Goltz, p. 75.) As a result, when the evaluation is situated within the confines of a really emotional debate, a quantitative, generalizable impact study can be something of a protection for the evaluator. The problem is that such a study may be neither feasible nor appropriate to the question. Further, even when it is the right study to do, if legislative policy staff are given a choice between timeliness and the kind of powerful methodology that could really be persuasive, staff will generally choose timeliness, so great are the political pressures of deadlines on the Hill. Thus, the translation from an evaluation question to a design that is appropriate (both methodologically and politically) is not always easy.

An eighth point regarding framework changes has to do with getting the agreement of the sponsor on the evaluation design, once it is completed. While this generally goes without saying in the executive branch, discussion of evaluation designs with legislative staff has been much less frequent. We have adopted this as a formal procedure in PEMD evaluations and we use the discussions about the design in several ways. First, the design will have come to some conclusions about how long the evaluation will take, what methods will be used, how likely we are to get answers to the questions posed, how unequivocal those answers are likely to be, how credible the study will be (especially if it is on a controversial subject), and what types and numbers of reports will be produced. The design meeting with the sponsor involves a staff

briefing, discussion of any changes the sponsor wants, and finally, agreement on how the evaluation will be conducted and reported. The issue of reporting relates once again to the problem of timeliness in the Congress: the design will choose among various types of products -- statement of fact, briefing report, testimony, full-length report -- based on the sponsor's timing needs.

A second way we have used the design discussions is to help initiate legislative staff into some technical aspects of evaluation. The necessity for this is dictated by misunderstandings that have arisen when congressional staff have not been carefully briefed on exactly what work is being pursued. The two biggest problems we've had have been overgreat expectations about what can be accomplished in a very short time, and unawareness of what the immaturity of a subject area can mean for the evaluation's ability to produce strong information. In one case, we were given only 12 months to produce a study that needed twice as much time. In another, we were expected to produce a methodology in a field where the paucity of research indicated that we could, at best, produce guidelines.

A third way in which the design meeting is useful is that congressional staff will typically explain to us how they expect to use our work, and this, of course, may well cause a modification in the design, or in the planned products. It also exposes us to important cumulative insights into the legislative process and into congressional relationships with the executive branch.

A ninth point on this third part of the continuum has to do with the prodigious importance for use that the overall credibility

of the evaluation will take on, and the kinds of efforts that have to be made to achieve that credibility during the conduct of the evaluation. Even though GAO already enjoys a remarkable reputation for institutional objectivity, we in PEMD ask independent researchers to look at our work on a quasi-continuing basis, so aware are we of the need for and benefits of strong critical review. As I've already noted, we often use panels in the conduct of a study. In addition, we ask different expert reviewers to look at our evaluation designs and draft reports. I have ongoing nightmares about testifying on the results of some evaluation synthesis we will have done, and hearing some Member of Congress say, "Well now, Mrs. Chelimsky, how does it happen that that seminal study, that crucial piece of work done by that very great researcher, Dr. X, from my home state, doesn't happen to figure in your synthesis?" At PEMD, we've done a lot of work and taken a great many pains -- mostly through the use of snowball sampling -- to ensure that we're never in the position of having overlooked one of those great, seminal studies. PEMD staff have somewhat different nightmares, mostly about releasing preliminary findings to data-hungry committees before they have been adequately referenced.

Here, the use of the framework conflicts once again with former evaluation practice. One of the Cronbach theses advised evaluators "to release findings piecemeal and informally to the audiences that need them" because "the impotence that comes with delay may be a greater risk than the possibility that early returns will be misread" (Cronbach, p. 6, thesis 47). A worse and much more dangerous possibility, I find, is not that those early returns

will be misread but that they'll actually be wrong, because they are early and have not been systematically checked. The rules at GAO about indexing, referencing, and quality assurance really stand us in good stead here. Staff who have not been a part of the evaluation team are the ones who do this work and during the sometimes incredibly detailed shakedown sessions about the support for this number, or the appropriateness of that data analysis, the cries and whimpers of maddened researchers can be heard throughout the division. The interesting thing is that, disagreeable though the quality control process usually is, nearly everyone is glad we have it. So we tend not to release information until we are sure it has satisfied the quality control process, because we have learned that a loss of credibility takes a long time to recover from and can in fact prove fatal. In sum, while we recognize absolutely the extreme importance of early information on the Hill, I would say that credibility is even more important: it is in fact the critical prerequisite for use over the long term. Whether the issue is fairness, balance, methodological quality, or just plain old accuracy, no effort to establish credibility is ever wasted. The memory of poor quality lingers long after the joy of meeting a deadline has been forgotten.

Part 4: The Translation of the Evaluation Findings Into Policy Answers that Respond to the Original Policy Question

Now we come to the fourth part of the continuum. This is where the evaluation results are translated into policy answers to the original policy question. Here what we are trying to do, of course, is communicate evaluation findings to a political audience

in such a way that the findings will be both well understood and persuasive. This again implies changes to typical evaluation practice and a tenth point I'd like to make speaks to the difficulty of establishing priorities among the findings. In politics, priorities are the heart of the matter: not ideology, not even consistency or persistency, but putting first things first at the right time. In evaluation, on the other hand, this is not only one of the most difficult things to achieve, it is rarely thought about at all. We evaluators have the habit of laying out all our results in the same careful, neutral phrases, leaving it to the scientifically trained audience to perceive the areas of success, of promise, of no-result, of failure. And we document everything, usually in the same neutral tone, with what has been described as "self-defeating thoroughness" (Cronbach, p. 6, thesis 49). The problem here is that all of the evaluation's findings seem important to its author. So, it is painfully difficult to order them in a policy context, to trim surgically what is not relevant, to condense, to rank, to decide not only which finding is the most important, perhaps, but which is the most important that policy can do something about.

Once again we hear the cries and laments of unhappy evaluators. But from the viewpoint of use, it is crucial to discriminate in communicating results between what is important and what is not, and to be clear about how the findings should or could be acted on.

Some findings, of course, are difficult to be clear about, except to say very bluntly that they are not clear. In other cases, we simply don't know the answer to the policy question.

It's important, I think, from a political viewpoint, for us to become comfortable with the idea of saying that a conclusive answer is not available when in fact it isn't. In our work, we have found that the knowledge that no knowledge exists is important policy information that decisionmakers do not necessarily have. In many cases, for example, the Congress will be told by an executive agency that there is no evidence of certain unfavorable results in some program. But what that often means is that the agency hasn't looked for such evidence. This is a very different thing from being able to say, We have looked seriously and systematically, and there is no evidence one way or the other. The point here is that if there is no evidence and this can be shown, then the way is cleared to get support for research that will, if possible, develop the needed information.

An eleventh point is that, in trying to make findings clear, useful, and effectively available to policymakers, communication techniques need to be developed which may depart quite largely from typical evaluation practice. Let me mention three here. A first effort has to do with banishing evaluation jargon. This is important, not just because policymakers don't understand it, but because it seems to them to hint at some private club, some "special" understanding, some incomprehensible ritual from which they are excluded. This gives rise not just to impatience but to a real irritation with the evaluator which carries over to the evaluation. It is the kind of irritation that Mencken expressed when he said a metaphysician was one who, when you remark that twice two makes four, demands to know what you mean by twice, what

by two, what by makes and what by four. The intelligibility of an evaluation report is critical for the use of the findings.

A second effort has to do with varying the written products so that the information can be presented briefly and succinctly, but is backed up with all the data support needed by a variety of audiences. A third effort involves paralleling the delivery of reports with as much oral communication as possible. This is the style of reporting that is most natural to policymakers, especially those on Capitol Hill. Congressional staff and Members of Congress are used to listening to constituents and lobbyists, they gather ideas and form judgments not only by reading but especially by talking with staff and colleagues, by holding committee hearings, by interacting personally with a wide variety of people. In this environment, oral briefings are crucial; testimony can be more useful than a major report. We now see video being pressed into service to help reconcile the tension between the need to document and the need for easy expression and understanding.

In general, when you think about the entire process I have been describing, starting with the original policy question, what has happened is that, in the evaluation, we have moved purposefully away from the anecdote or war story as a credible answer to a policy question. As evaluators, we're typically trying to get a sense of the size of a problem, of its range, its frequency, its direction, its average characteristics. But after having determined the evaluation results, when we must move back into the political forum the policy question emerged from originally, we find that one of the most effective ways to present findings is to

rediscover the anecdote, to illustrate the general findings via specific cases and analogies that graphically focus attention on and explain the larger points. This is, of course, a totally different use of anecdotes. Here they are not asked to stand alone as answers to policy questions but instead, are used as representatives of the broader evaluative evidence, to explain the findings, to increase the political understanding and comfort-level, and, above all, to improve the likelihood that the evaluation findings will be used.

Part 5: Use of the Evaluation Findings and the Generation of New Policy Questions

Now we come to the buckling-the-buckle or fifth and last part of the continuum, which involves use and the generation of new questions. In traditional agency-focused evaluation practice, it has often been quite difficult to trace the use of evaluation findings. This is less true with legislative evaluation, where use can be measured in a number of ways. We have seen our own evaluations figure in program reauthorization (one example is our assessment of homeless and runaway youth centers which was a major factor in the doubling of the centers' reauthorization funds; another example is the change in the law -- based on our findings about the situation of working mothers in the AFDC program -- that was enacted to allow a prolonged period of health insurance for earner families). Hearings are often organized around our findings (examples are recent hearings on such issues as the Medicare Prospective Payment System, the defense industrial base, student aid, educational research, hazardous waste, poverty indicators,

drinking age laws and many others). Our findings and recommendations are incorporated in legislative direction to the executive branch (an example is our finding in the science and math teachers' study of the need for evaluation in NSF teacher-training programs, which NSF was then directed to perform by the Senate Appropriations Committee). And finally, we can measure use of our findings in legislative debates that directly affect up-or-down congressional votes (here I would cite our work on chemical warfare which, year after year, figures notably in the national argument about whether or not to produce binary chemical weapons). So my twelfth point about framework differences when legislative evaluation is included, is that, in the Congress, use occurs -- indeed, use is the rule, not the exception -- and it is measurable.

In addition, we achieve a further use of evaluation findings through the synthesis method we have developed which brings together previous evaluations in a topic area, analyzes and compares them -- re-uses them, in sum -- with the intent of developing a general finding about knowledge in that area. We will be engaging in the same process of re-use as we develop the new evaluation planning review that I mentioned earlier.

A thirteenth point concerns the evaluator's ability to influence successive rounds of policy questions that need to be asked but that -- for one reason or another, as I noted earlier -- may not be generated by policy actors. I see two types of questions that we must try to influence. The first type is descriptive: it is truly amazing how difficult it is to persuade policymakers of the importance of descriptive information. Our

work on hazardous waste found that we simply do not know how many million metric tons of the stuff are being generated every year in this country. Questions like, What is the size of a problem, how many people are involved in a program, what characteristics they have, what services they receive and how they feel about them, or in a technological communications program, say, how many messages are sent, who sends them and who receives them, these types of questions just do not get asked in some agencies and, as a result, more sophisticated questions about program benefits or effects cannot be adequately answered when the Congress needs those answers.

One political factor that contributes to the lack of strong program description is that "the more political support there is for a program, the more limited will be the available systematic information on that program" (Stromsdorfer, p. 6). Defense programs are a good example of this, and the fact that much of the information involved is classified, non-public information doesn't help. Rarely is there archiving of test and evaluation findings, even on a need-to-know basis, and so every effort to evaluate every weapon system has to start from scratch. The Congress is, however, beginning to look a little more carefully at the evaluative basis for moving into production with some weapon systems, so this may change.

On the other hand, relatively unpopular programs tend to produce more information and to be evaluated more often. Herman Leonard has just published a new book in which he looks at some popular and unpopular programs and their ability to spend money

without drawing attention to themselves (Leonard, 1986). He points out, among other things, that Social Security has been able to commit over $100 billion annually for the last 40 years without much discussion, while political controversy surrounds AFDC's $10 billion or so annual budget. What this means for evaluation is that a large amount of systematic descriptive information is available on AFDC while a host of other more popular programs (like defense and social security) have been largely able to avoid much public scrutiny.

The second type of question that doesn't get asked much in executive agencies (it does most definitely get asked in the Congress) is the impact question. Here the impediment remains twofold: the threat to decisionmakers and the paucity of incentives for accountability (Chelimsky, 1977, Vol. II, p. 36-39). But whereas, over the years, we did see quite a few of these studies done in the executive branch despite the threat they unquestionably pose, over the past 5 years, we have seen a marked -- very marked -- reduction in results-oriented evaluations. So it is quite clear that if these evaluations are to be done in the executive branch under this administration, I think the Congress will have to mandate them, as has often happened in the past. Indeed, a complicating problem is that, not only are the evaluations not being done, the administration has been phasing out much of the data system infrastructure that is needed for the conduct of these studies.

From the viewpoint of framework changes, what this means is that while it is true we have given up, for policy studies, some of

our traditional freedom to choose programs and questions, leaving these now to emerge from the political arena, it is not clear we would have been very successful in any case in getting agency policymakers to ask many impact or descriptive questions. That is, where sponsored policy studies are concerned, it may be that the evaluator's freedom to ask certain questions has always been more nominal than real. On the other hand, by noting the absence of answers to these questions in our findings and by recommending to the Congress that these questions be asked, it is possible we can get more attention paid to their importance the second-time around.

What Have We Learned?

In summary, then, what do I think we have learned about the politics of program evaluation? I feel I can point to a number of lessons.

(1) I think we've learned that, to be successful, we must be useful to others than ourselves. That means understanding the political system in which evaluation operates, and understanding the information needs of those policy actors who are evaluation's users.

(2) We've learned to conceptualize the political system a little more broadly, to include all sectors which make the kinds of policy decisions into which evaluation feeds (that is, not just executive but also legislative branch policymaking). Indeed, if the judicial branch should continue increasing its use of cost/benefit and risk analysis, we will need to study and include the judiciary as well. Such study would certainly include evaluative ramifications on the kinds of evidence that need to be

produced, and political ramifications of judicial use on decision-making in the executive and legislative branches. This is important because, as we have seen, evaluation's interactions with legislative and executive branch policy actors can be profoundly modified by the inclusion of a new set of policy actors from a different branch of government. This is because policy is never made in isolation: legislative policy affects executive branch policy integrally, and the converse is equally true. Evaluators need to understand political interactions, no matter which branch they are working for.

(3) We've learned that evaluators working for the legislative branch will be pressed to go beyond earlier social program horizons; the congressional functions require evaluative skills in such areas as defense, the environment, energy and natural resources, as well as in health, education and welfare, for example.

(4) Planning for use involves understanding the political conditions under which evaluation findings will be used. We've learned that two of these conditions are that (a) the policy question posed must be of fundamental interest to the intended user, and (b) that the eventual evaluation findings must answer that question. Both of these conditions empower the decisionmaker at the expense of a certain amount of evaluator discretion.

(5) We've learned that a third political condition for use is credibility. If the evaluation is assailable on grounds of poor methodological quality or on grounds of partisanship, this will reduce the use made of its findings in the Congress. Put another

way, the high quality and acknowledged objectivity of a study cannot ensure use if the policy question is not of interest; but poor quality and subjectivity will weaken use, no matter what the question is.

(6) Still a fourth condition for use is timeliness. Evaluators working for the legislative branch must be extremely concerned about the timing of the final product and how it dovetails with congressional policy cycles and plans for use. We've learned that what is most important sometimes is not having the best design, but having an adequate design that will bring the findings in at the time they were promised.

(7) Overall, we've learned that there is no perfect evaluation design but that, rather, evaluators must try to achieve a balance involving timing, methodological strength, and cost. Often it's a matter of degree: a methodology that doubles the time necessary for performing an evaluation will be unacceptable no matter what its advantages. But one which increases the time moderately while strengthening the conclusiveness of the information to be obtained may be enthusiastically accepted under some political circumstances.

(8) Planning for use always incorporates the question, who will implement the study recommendations? We have learned that good relations with the implementing executive agency may be the sine qua non of that ultimate use.

(9) Evaluators have not conducted many prospective studies. But we've learned that one of the most felicitous times for the introduction of evaluative thinking into the policy forum is before

a new bill is proposed. New techniques need to be developed to enable doing this work on a very short time-schedule.

(10) Panels of experts can serve a very important legitimizing function. We've learned that carefully chosen panels are a powerful aid in strengthening the political crediblity of an evaluation's findings, especially when the study is in a controversial area.

(11) It's important for an evaluator to have mastered a wide variety of evaluation approaches. No single design or method can be counted on to address the diversity of policy questions that political debates engender.

(12) The transition from policy question to evaluative question to evaluation design is not a straightforward process and usually involves several iterations. We have learned that, no matter how many negotiations there have been, it is wise to be sure the user agrees with the finished design and understands both how the study will be conducted and what kind of information will be available at its conclusion.

(13) We have learned that indexing, referencing and general checking of draft reports are activities that are worth their weight in gold; although time is important, and all this checking does have to be planned for, there is enormous risk in any weakening of credibility.

(14) Prioritization of the evaluation findings is also a condition for use. We have learned that telling all is tantamount to telling nothing. The important thing is to answer the policy question as clearly and simply as possible, to emphasize a few

critical and striking numbers, and to do all that in such a way as to highlight those findings that give rise to policy action. Quite a lot of time needs to be devoted to thinking out how we will present our findings so that they can be intelligible to our several audiences.

(15) In the legislative branch, the use of evaluation findings is the rule, not the exception. In these past 6 years of working with the Congress, we have learned that evaluations are a demonstrably important adjunct of policymaking, that they do figure in go/no go decisions, and that they often set the agenda for national debates.

(16) A further use of evaluations can be achieved via studies employing meta-analysis or evaluation synthesis. Such secondary evaluation is a major tool in helping decisionmakers make sense out of conflicting findings and improve the quality of information available to them.

(17) Descriptive questions are not typically raised by policymakers, which means that systematic information about some programs is often missing. We have learned that emphasizing in an evaluation report that such information is indeed both missing and needed can help to get it provided at a later date.

(18) Impact questions seem to be asked less and less by executive branch agencies, more and more by legislative policy-makers. We have learned that a legislative user can have very strong leverage on programs, through the authorization and appropriations process or through oversight.

(19) Finally, the fact of taking political processes into account when doing an evaluation transforms the way in which time is allocated. We have learned to devote much more time to negotiation, to discussion, briefing, accuracy-checking, and to prioritizing and presentation than we used to do.

Overall, I would sum up all of these political lessons learned by saying that things have changed insofar as evaluation use is concerned. Beyond any doubt, the use in policymaking that we're seeing now by the Congress is real, measurable and growing dramatically. But if we <u>are</u> making progress, we owe a large debt of gratitude to those who cried failure in the past, because they focused our attention on the political environment in which we expected to be useful but knew very little about. This was important because use doesn't happen by itself. Instead it requires some major modifications of traditional evaluation practice, such as those I described earlier.

This is not to say that it's time to cry victory. We can't do that so long as the administration continues to cut back its results-oriented evaluation work, although we <u>can</u> hope that executive agencies may soon begin to realize there is growing risk in failing to evaluate their programs: if they don't, legislative agencies will.

On the other hand, I think we <u>have</u> dispelled the notion that evaluation isn't useful in policymaking. So there is every reason to be hopeful about the future of evaluation research and practice, given the important role it is playing in improving the quality of information available to decisionmakers.

References

Ilene N. Bernstein and Howard E. Freeman, Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Studies, New York, Russell Sage Foundation, 1975.

Eleanor Chelimsky, Program Evaluation: Patterns and Directions, American Society for Public Administration, 1985, Washington, D.C.

Eleanor Chelimsky, Proceedings of a Symposium on the Use of Evaluation by Federal Agencies, Vol. I, MITRE Corporation, 1977.

Eleanor Chelimsky, An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies, Vol. II, MITRE Corporation, 1977.

Eleanor Chelimsky, "The Politics of Program Evaluation," in Sage Yearbook in Politics and Public Policy, ed. Dennis Palumbo, Sage, in press.

Lee Cronbach and Associates, Toward Reform of Program Evaluation, Jossey-Bass, 1980.

David Easton, "An Approach to the Analysis of Political Systems." in World Politics, Vol. IX, April 1957, Princeton University.

David H. Florio, Michael M. Behrmann and Diane L. Goltz, "What Do Policy Makers Think of Educational Research and Evaluation? Or Do They?" Educational Evaluation and Policy Analysis, November-December 1979.

Herman Leonard, Checks Unbalanced: The Quiet Side of Public Spending, Harvard Press, Cambridge, Massachusetts, 1986.

Charles E. Lindblom, The Policy Making Process, Englewood Cliffs, New Jersey, Prentice Hall, 1968.

Ernest W. Stromsdorfer, Social Science Analysis and the Formulation of Public Policy: What the President "Knows" and "How" He Comes to "Know" It, Conference on Social Experimentation, National Bureau of Economic Research, March 1981.

James L. Sundquist, "Research Brokerage: The Weak Link" in Knowledge and Policy: The Uncertain Connection, National Research Council, 1978.

U.S. General Accounting Office, Designing Evaluations, Program Evaluation and Methodology Division Methodology Transfer Paper 4, July 1984.

U.S. General Accounting Office, "Teenage Pregnancy: 500,000 Births a Year but Few Tested Programs," GAO/PEMD-86-16BR, July 21, 1986.

Carol H. Weiss (1979) in "The Many Meanings of Research Utilization" in Eleanor Chelimsky, Program Evaluation: Patterns and Directions, American Society for Public Administration, 1985, Washington, D.C.

Carol H. Weiss, "Where Politics and Evaluation Research Meet," in Evaluation, 1:3/1973.